Nouf Aljohany

220058009

An Introduction to Speech Systems

Speech recognition has been defined to mean the process of converting an audio or acoustic signal that has been captured in some fashion into a set of words that can be read or visually recognized. In short, it is converting a signal for the ears into a signal for the eyes.

Speech recognition systems are used to transcribe an audio track in the same language or they can be used to translate from one language to another. This speech-to-text and machine translation process has special requirements that must be considered. Issues and challenges include accuracy, punctuation, sentence boundaries, tokenization, and numbers and dates processed into correct written form.

Voice recognition systems are used with computer systems to recognize and record spoken words. The comprehension by the system and further actions are not implied in this definition. Many available voice recognition systems can work with many thousands of words. These systems have difficulty with different speakers, accents, and pronunciations that are used. Speaking patterns vary from person to person and systems often have difficulties correctly recognizing different speech patterns from different people. One beneficial application of these systems that is often mentioned is that they are useful for handicapped people who are unable to use a keyboard or a mouse, but are capable of giving verbal commands. These sophisticated systems are often quite expensive but as technology develops the cost decreases.

<http://www.abilitynet.org.uk/content/factsheets/pdfs/Voice%20Recognition%20Software%20-%20An%20Introduction.pdf>

Text-to-speech systems are used to reverse the process. These take textual information and convert it to an audio sound medium. An example of this is *spokentext.net* which records English, French, Spanish or German language files such as PDF, Word, plain text, PowerPoint, and web pages, and convert them to speech automatically. This tool creates audio recordings in the four languages of any text content.

<http://www.laits.utexas.edu/hebrew/personal/tts/table.html>

An example of a speech-to-text online system is the YouTube technology in which automated spiders "listen" and transcribe words contained in the audio track of a video.  This transcription, in the form of metadata, is then put into the content of the video. This new speech-to-text technology was used in the 2008 US presidential campaign for videos of both of the candidates. This text can then be searched on in the YouTube search facility.

<http://www.beet.tv/2008/07/youtube-has-voi.html>

Arabic enabled speech recognition software is also being developed. An example of this is Sakhr Software that provides speech recognition and text-to-speech technology for application development. This commercial software is designed to understand natural Arabic speech in different Arabic accents. This software was initially developed for telephone applications.

<http://international.sakhr.com/arabic-speech-recognition-and-arabic-TTS.html>